DOCUMENT RESUME

ED 417 722 IR 056 947

AUTHOR Ostrow, Stephen E.
TITLE Digitizing Historical Pictorial Collections for the Internet.
INSTITUTION Council on Library Resources, Inc., Washington, DC.
ISBN ISBN-1-887334-57-2
PUB DATE 1998-02-00
NOTE 44p.
AVAILABLE FROM Council on Library and Information Resources, 1755 Massachusetts Ave., N.W., Washington, DC 20036 ($20).
PUB TYPE Reports - Descriptive (141)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Access to Information; *Internet; *Library Automation; *Library Collection Development; Library Material Selection; *Library Technical Processes; Nonprint Media; *Photographs
IDENTIFIERS *Digital Collections; *Digital Technology

ABSTRACT

This report is one in a series from the Council on Library and Information Resources that explores the issues and problems that institutions must address when undertaking digital conversion projects. The focus of this report is on projects that mean to provide widespread access over the Internet to large historical pictorial collections, using an expandable base of digital files on a server. Topics addressed include: how the use of digitized surrogates can differ from, as well as complement, the use of original primary resource materials; criteria for selecting items and collections for digital conversion; associated issues raised by providing digitized access over the Internet, including fair use, reproduction rights, and the creation of bibliographical records for digital images; and implications for preservation. Contents: (1) The Nature and Use of Historical Pictorial Collections; (2) Observations about Images in a Digital Environment; (3) Digitizing Historical Pictorial Collections; (4) Beyond Images: Rights and Records; (5) What to Digitize: The Questions to Ask; (6) Conclusion. Appended are the Library of Congress, NDLP Project Planning Checklist and the Copyright and Other Restrictions Which Apply to Publication and Other Forms of Distribution of Images: Sources for Information. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 417 722

Digitizing Historical Pictorial Collections for the Internet

by Stephen E. Ostrow
February 1998

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

M. K. Sitts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Council on Library and Information Resources

Commission on Preservation and Access

Digital Libraries

Economics of Information

Leadership

IR056947

BEST COPY AVAILABLE

2

Digitizing Historical Pictorial Collections for the Internet

by
Stephen E. Ostrow

Council on Library and Information Resources, Washington, D.C.

February 1998

About the Author

From 1984 until his retirement on December 31, 1996, Dr. Stephen E. Ostrow was chief of the Prints and Photographs Division (P&P) of the Library of Congress. At the time of his retirement, more than 220,000 of P&P's images were available electronically over the Internet and/or in the division's reading room, with 590,000 more in process. Prior to joining the Library of Congress, Dr. Ostrow taught art history at the University of Missouri and Brown University, was chief curator and then director of the Museum of Art, Rhode Island School of Design, and served as dean of the School of Fine Arts at the University of Southern California. He received his bachelor's degree in fine arts from Oberlin College, and his master's and doctor's degrees in art history from the Institute of Fine Arts of New York University.

Commission on Preservation and Access

The Commission on Preservation and Access, a program of the Council on Library and Information Resources, supports the efforts of libraries and archives to save endangered portions of their paper-based collections and to meet the new preservation challenges of the digital environment. Working with institutions around the world, the Commission disseminates knowledge of best preservation practices and promotes a coordinated approach to preservation activity.

Digital Libraries

The Digital Libraries program of the Council on Library and Information Resources is committed to helping libraries of all types and sizes understand the far-reaching implications of digitization. To that end, CLIR supports projects and publications whose purpose is to build confidence in, and increase understanding of, the digital component that libraries are now adding to their traditional print holdings.

ISBN 1-887334-57-2

Published by:

Council on Library and Information Resources
1755 Massachusetts Avenue, N.W., Suite 500
Washington, D.C. 20036

Additional copies are available for $20.00 from the above address. Orders must be prepaid, with checks made payable to the Council on Library and Information Resources.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences - Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

Copyright 1998 by the Council on Library and Information Resources. No part of this publication may be reproduced or transcribed in any form without permission of the publisher. Requests for reproduction for noncommercial purposes, including educational advancement, private study, or research will be granted. Full credit must be given to both the author and the Council on Library and Information Resources.

Introduction

This report is one in a series from the Council on Library and Information Resources that explores the issues and problems that institutions must address when undertaking digital conversion projects. The focus of this report is on projects that mean to provide widespread access over the Internet to large historical pictorial collections, using an expandable base of digital files on a server. The report does not deal with CD-ROMs distributed in multiple copies as a digital publication, or digital images as an alternative to microfilms and reproductions in print publications *per se*.

Large-scale digitization programs are complex to plan and execute, and they have far-reaching implications that often go well beyond the staff directly involved in project management and implementation. Stephen Ostrow has a uniquely informed perspective on the impact of digital conversion projects on an institution's collections, staff, and even mission. As former chief of the Prints and Photographs Division of the Library of Congress (1984-96), he grounds his approach not in theory but in practice. He experienced the strict discipline of maintaining core services—acquisition, preservation, processing and cataloging, documentation, and public access—while scaling up what was initially the provision of analog images in the reading room to a full-blown digital conversion effort meant for dissemination over the Internet. His experience, therefore, leads him to focus on the various ways that this rapidly evolving technology of access affects traditional services and values. This report, which reflects the specifics of one institution—its collections, organization, history, and mission—nonetheless offers important analogies with many other repositories of historical pictorial materials.

The report focuses on a number of considerations important to institutional planning, including:

- how the use of digitized surrogates can differ from, as well as complement, the use of original primary resource materials;
- criteria for selecting items and collections for digital conversion;
- associated issues raised by providing digitized access over the Internet, including fair use, reproduction rights, and the creation of bibliographical records for digital images;
- implications for preservation.

The author concludes that digital technology is moving libraries in dramatic new directions. Compared with the *evolution* from catalog cards to machine-readable records, moving from original visual documents to electronic surrogates falls in the realm of a *revolution*. The promise of the technology is that, although digital access is not a replacement for access to the original, it nonetheless can provide significant new service options to a wide variety of users that go well

beyond the traditional scholarly audience. The questions for institutions with large pictorial collections are both practical and philosophical. They must determine how best to manage the incorporation of this new option for access and service into their strategic planning for reader services, bibliographic control, and preservation, and they must do so in ways that relate to an institutional mission that is being challenged by the ability of digital technology to reach new audiences.

Abby Smith
Preservation and Access Program Officer

Contents

Author's Acknowledgements **vi**

I: The Nature and Use of Historical Pictorial Collections **1**
- The Nature of Collections **1**
- Uses of Historical Pictorial Collections **6**
- Organization and Service of Collections **8**
- Preservation Issues **9**
- Digital Images and the Reading Room Experience **9**

II: Observations about Images in a Digital Environment **12**
- The Digital Infrastructure **13**
- Purpose, Audience, and Resolution **13**
- The Pace of Change **16**

III: Digitizing Historical Pictorial Collections **17**
- Why Create Digital Images? **17**
- Primary and Secondary Education **18**
- Retrieval and Service **19**
- Digital- and Film-Based Preservation Surrogates **20**

IV: Beyond Images: Rights and Records **23**
- Rights **23**
- Records **25**

V: What to Digitize: The Questions to Ask **26**

VI: Conclusion **28**

Appendix A: Library of Congress, NDLP Project Planning Checklist

Appendix B: Library of Congress, Copyright and Other Restrictions Which Apply to Publication and Other Forms of Distribution of Images: Sources for Information

All photos in this report are reproduced from the Prints and Photographs Division, Library of Congress, Washington, D.C.

Author's Acknowledgments

I am indebted to my colleagues at the Library of Congress for their frequent and enlightening discussions over the years of many of the topics addressed in this paper. My special thanks go to Elisabeth Parker, who served with me as the assistant chief, Mary Ison, head of Reference, Bernard Reilly, head of the Curatorial Section, and Helena Zinkham, head of the Technical Services Section of the Prints and Photographs Division. While I take full responsibility for the opinions and conclusions expressed in this essay, the fact remains that many of my views developed in the context of my colleagues' contributions. This does not imply that they would agree with what I have written. Additionally, I am indebted to Carl Fleischhauer, technical coordinator for Digital Conversion of the National Digital Library Program at the Library of Congress, for having nurtured so patiently my knowledge and understanding of the technical aspects of the subject.

Stephen Ostrow

I: The Nature and Use of Historical Pictorial Collections

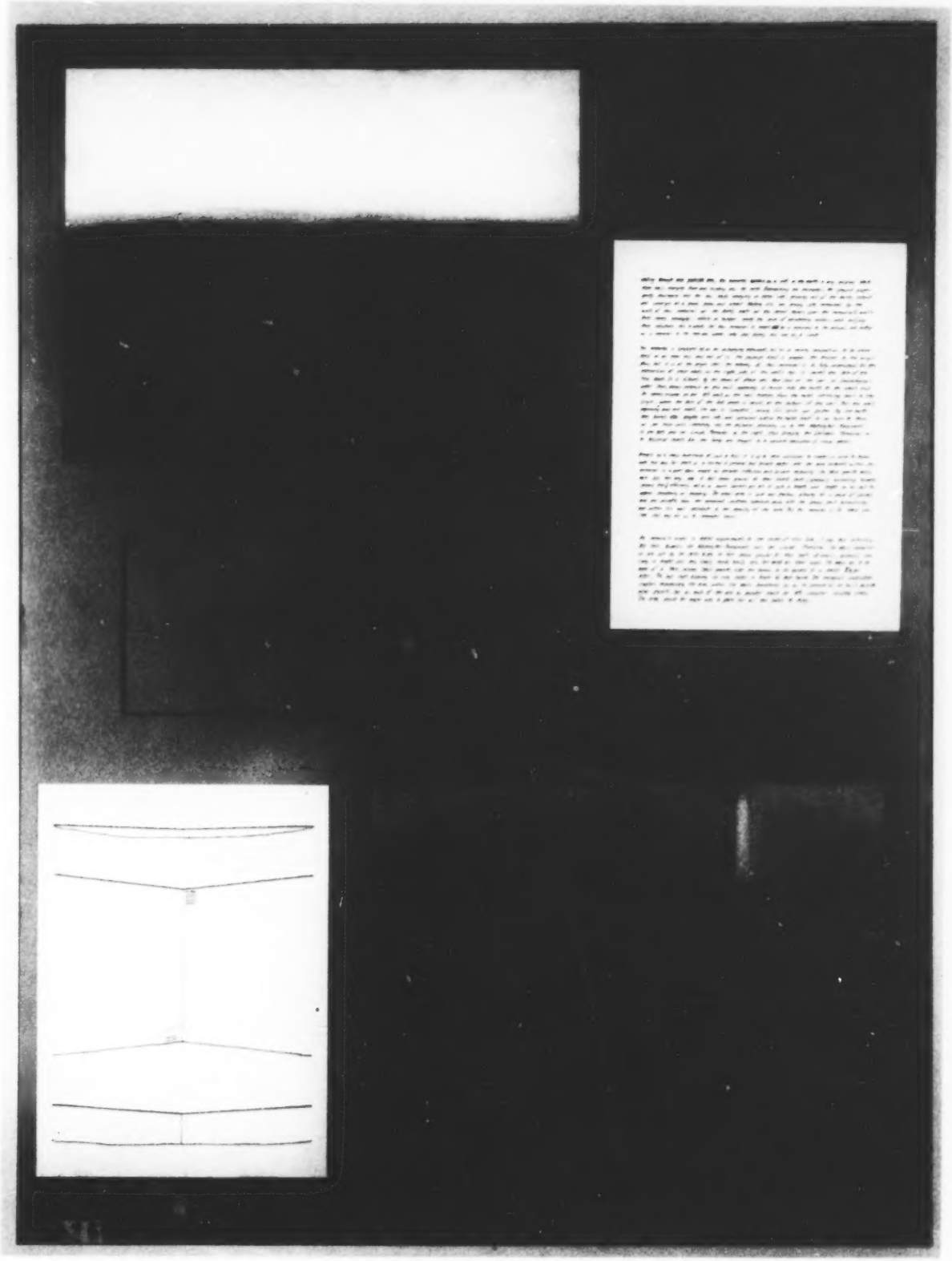
> *There is agreement among academics, educators, and librarians alike about the rationale for creating ... a national digital library to provide wider electronic access to knowledge and information. Such access would (1) vastly enhance education, scholarship, arts, and science; (2) preserve and make accessible unique primary materials about our heritage; (3) make America more competitive in world markets; (4) foster an informed citizenry.*
>
> *The National Digital Library Program—A Library for All Americans*, The Library of Congress, February 1995, p. 5

Large historical pictorial collections present special problems and considerations in the process of selecting for digitization. Until recently, when an institution made a decision about the acquisition of pictorial materials, it would consider such things as their content, format, physical condition, and relation to extant holdings, based on traditional and well-known assumptions about their potential use in research. All subsequent decisions about their preservation, bibliographical control, and access were made within the context of use in the reading room and, very often, additional use for reproduction in print or broadcast. A review of the nature, organization, and preservation of large historical pictorial collections and the ways they are used traditionally in reading rooms will help set the stage for our discussion of both the limitations of digital surrogates and the constructive role that they can play in research.

The Nature of Collections

Historical pictorial collections can range from thousands to millions of images, most of which are originals and many of which are also of unique artifactual value. Most collections, encompassing photographs and works on paper, include documentary photographs, fine and popular prints and drawings, posters, and architectural and engineering drawings. Among the many media included are glass-plate and film negatives; daguerreotypes, ambrotypes, and tintypes; silver gelatin prints and color transparencies; lithographs, woodcuts, and etchings; and pen-and-ink, watercolor, and chalk drawings.

It is their nature as evidence of things past and documentary sources for history, and not necessarily their aesthetic value, that makes these collections valuable to researchers. Photographs provide the look of things that have long since disappeared. Architectural drawings document stages in the creative process. Political cartoons portray transient attitudes and opinions that have become arcane. Moreover, researchers can inspect the original documents, using the artifactual evidence to confirm their legitimacy as historical



BEST COPY AVAILABLE

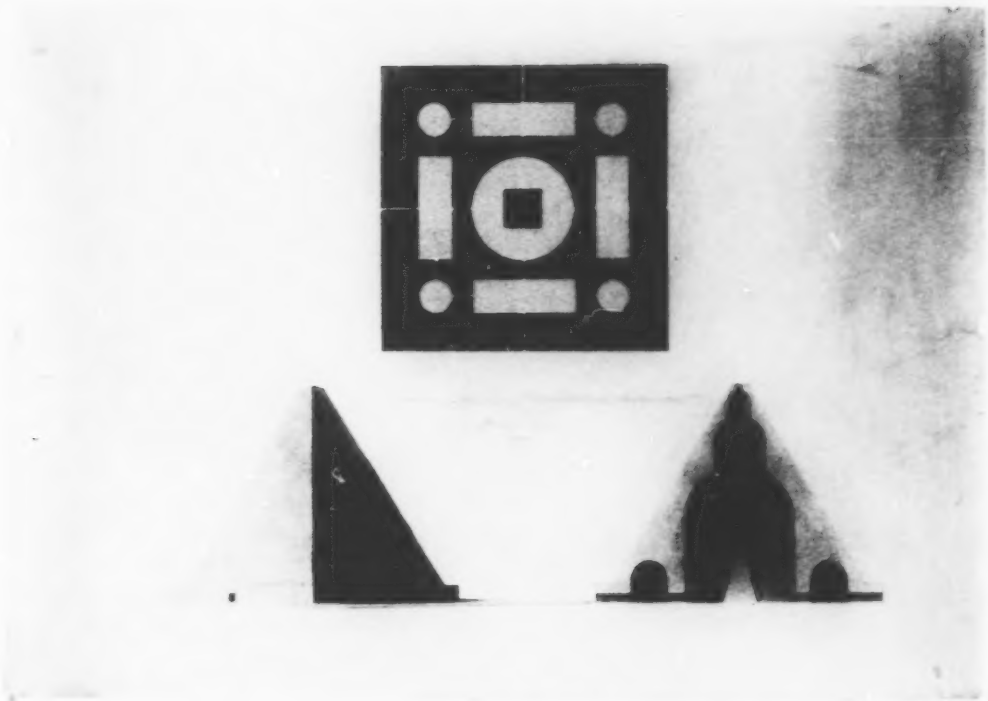
texts. Unlike many museum departments with pictorial collections, prints and photographs divisions in research libraries and archives tend to have collections that have not been selected by connoisseurs for aesthetic reasons or because of their artifactual importance, or edited by curators for purposes of exemplification. On the contrary, they are gathered comprehensively, to the extent possible, and it is the very completeness of the record that makes it a valuable source of evidence. Even items treasured individually as artifacts—for example, the earliest known daguerreotype portrait of Abraham Lincoln, the earliest known presidential campaign poster, a vintage photograph of the Manzanar internment camp by Ansel Adams, or a vintage photograph of the London blitz by Toni Frissell—are enhanced significantly if they are seen in the context of a larger collection of linked items.[1] This archival collecting practice, so distinctive to the acquisitions policies of research libraries, archives, and historical societies, lends to collections as such an intrinsic value for researchers, who come to rely on the comprehensive nature of the record as part of doing research in those institutions. A few examples of how these policies have shaped actual collections and, in turn, research strategies will illustrate this point.

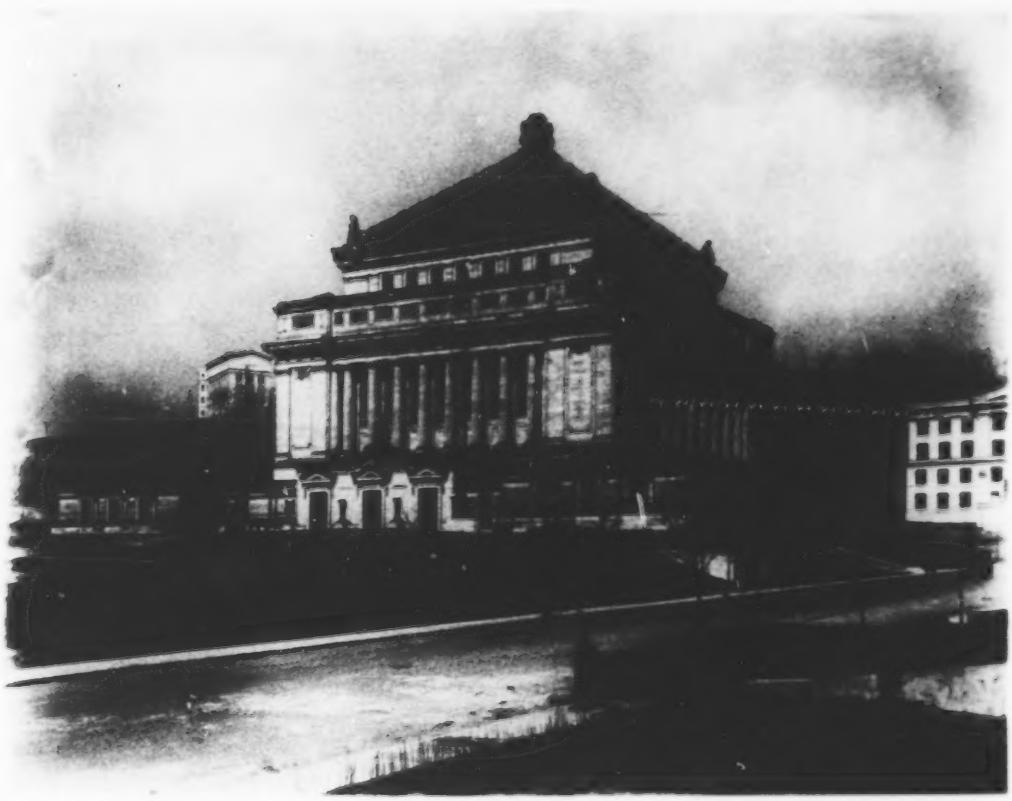
Maya Lin's presentation panel for the 1981 Vietnam Veterans Memorial design competition, with its eloquent drawings that suggest rather than describe the final monument, is part of a collection at the Library of Congress that documents almost all of the competition for that memorial. **(Fig. 1)** The panel itself has enormous authority as an artifact and tells us much about the genesis of this seminal monument when the artist first formalized the idea. When the panel is put alongside the design boards and color slides from the more than 1,400 submissions, one is struck by the wide range of choices that was available to the jurors. That they chose this groundbreaking design by a then-unknown student is quite remarkable. Both the process of design by an individual and the process of decision by a jury are open for close scrutiny when the entire competition archives is at hand.

←
[Fig. 1] Maya Ying Lin, *Vietnam Veterans Memorial Competition*. Presentation panel in mixed media on paper. 1981. Gift of the Vietnam Veterans Memorial Committee. Original in color. (LC-USZ62-90338)

This collection of submissions in turn forms one segment of a compilation of more than 40,000 architecture and design drawings related to the greater Washington area that are available to researchers who visit the reading room in the Library of Congress. It contains designs for other monuments and memorials, both built and unbuilt, such as Peter Force's competition drawing for the Washington Monument that dates from 1837. **(Fig. 2)** These additional materials can further delineate the innovative character of Maya Lin's design by allowing examination of the extensive history of such monuments in the nation's capital and its immediate surround. To take the study one step further, the collection of about 28,000 photo-

1. All specific examples cited in this essay are in the custody of the Prints and Photographs Division of the Library of Congress. For an overview of these collections and the research areas that they inform, see *Library of Congress Prints and Photographs: An Illustrated Guide*, Library of Congress, Washington, D.C., 1995. The *Guide* also is available on the Internet (http://lcweb.loc.gov/coll/print/guide).





BEST COPY AVAILABLE

graphic images from the Detroit Publishing Company broadens the reference to the Eastern United States with such monuments as Henry Hornbostel's *Soldier's Memorial* in Pittsburgh, the winning design of a 1907 competition. **(Fig. 3)**[2] While no one researcher is likely to pore through each and every one of the images available in the reading room, having such extensive holdings on related topics ready-to-hand does in fact broaden research horizons and make possible that "serendipity" of discovery so often alluded to by seasoned scholars.

In the case of photographic morgues, for example, one can even speak of the content of the image being defined primarily as part of a larger whole. News magazines take many photographs to document an event; usually only a few are chosen to appear in print. The value of all the others, and the very reason for creating a morgue and saving all the unpublished images in the first place, is that, en masse, these images tell a story that a single image cannot. An example from the *U. S. News & World Report* collection, is the series of photographs and contact sheets documenting the October 15, 1969 Peace Rally, in Washington, D.C. The richness of texture, the accretion of detail, and the shifting perspectives of the three staff photographers as they snap one frame after another in all directions create, in the totality of the images, an incomparable witness to history.[3] The multiple images reestablish that event and concretely introduce the passage of time, the variety of locations, and the diversity of the participants. They document students in windbreakers and jeans, arms linked with businessmen in suits and ties; a candlelight vigil at the Washington Monument, a rally at Selective Service Headquarters, and a march to the White House; and signs that range from the verbose ("BEM Business Executives Move For Vietnam Peace" and "Down With Dirty Air, Dirty Water, and Dirty Wars") to the succinct ("Professionals For Peace" and "Out! Now!"). **(Figs. 4-5)**

Another example comes from the social historian, John Vlach, who provides testimony of how he used photographs and drawings in the Historic American Buildings Survey (HABS) as primary documentation when he was researching his book *Back of the Big House*. His stated purpose was to describe "the architectural settings of plantation slavery and . . . some of the ways in which black people may have transformed those architectural settings into places that best served their social needs."[4] Although he also relied on written records and oral histories, Vlach asserts that HABS visual documentation was central to his research, in part because of the sheer numbers of slave structures and surrounding spaces that were drawn and

← [Fig. 2] Peter Force. *Architectural drawing for the Washington Monument, Washington, D.C.* Competition drawing showing pyramidal monument as plan, elevation, and section in wash, ink, and graphite on paper. 1837. (LC-USZ62-114247)

← [Fig. 3] Henry Hornstobel, architect. *Allegheny Co. Soldiers' Memorial, Pittsburgh, Pa.* Winner of a 1907 competition. Opened in 1911. Black and white glass plate negative. Detroit Publishing Company. Gift of the State Historical Society of Colorado. (LC-D4-72231)

[2] Records of the Architectural Design & Engineering Drawings, and records and images of the Detroit Publishing Company, are available on the Internet via the Prints and Photographs Division Online Catalog (http://rs6.loc.gov/pp/pphome.html).

[3] There are ten contact sheets containing about 270 images by the newspaper's staff photographers, Warren K. Lefler, Thomas J. O'Sullivan, and Marion S. Trikoso (folder numbers 21635-21637, 21648-21653).

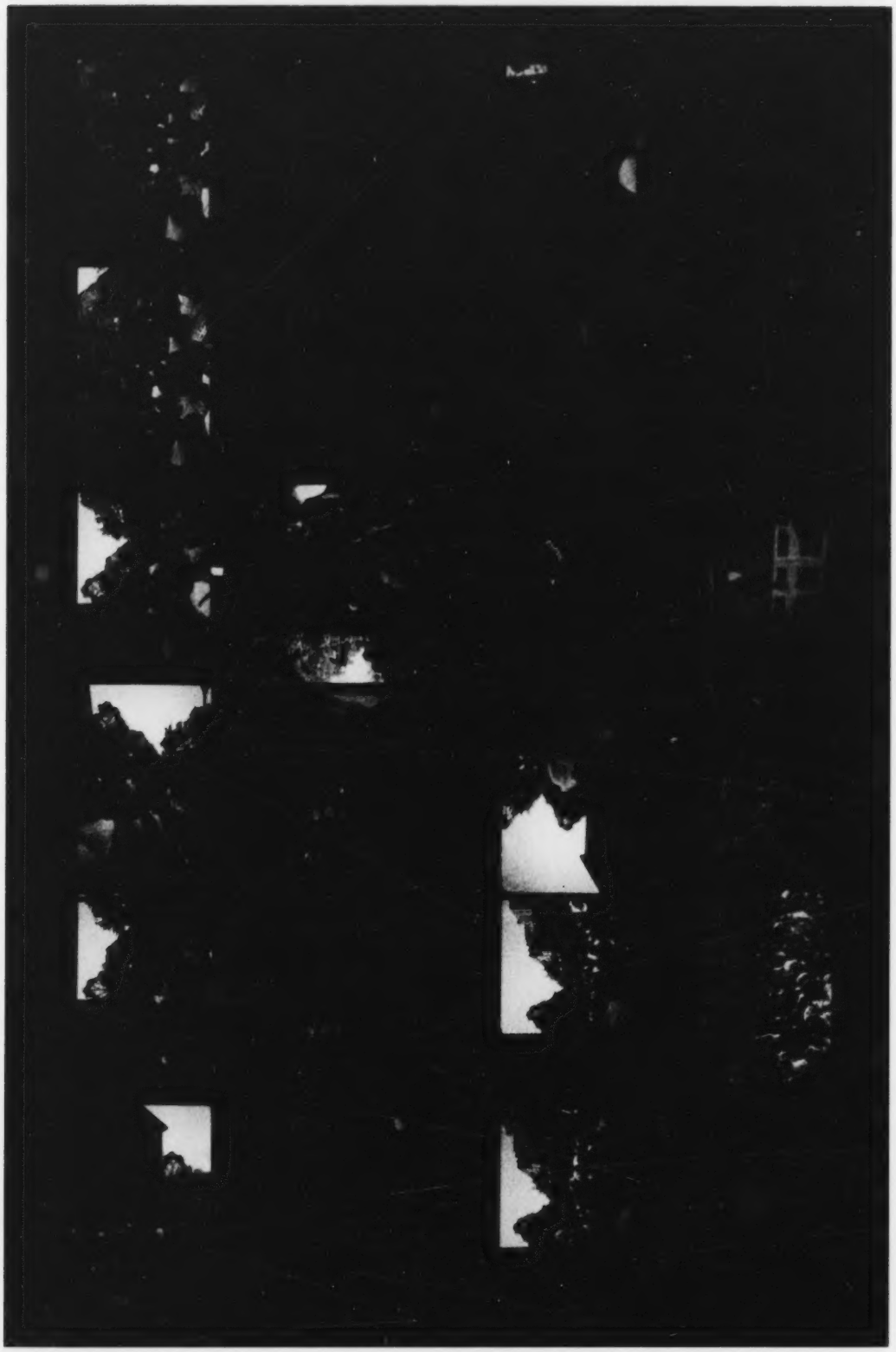
[4] John Vlach, *Back of the Big House: the Architecture of Plantation Slavery*, (Chapel Hill, N.C.: University of North Carolina, 1993), Preface, p. x.

photographed, in part because of the high quality of the measured drawings and photographs.

↑ →
[Figs. 4,5] Warren K. Lefler, Thomas J. O'Halloran, Marion S. Trikoso. *Peace March , Washington, D. C., October 15, 1969.* U.S. News & World Report Magazine Collection (LC-U9-21637)

Uses of Historical Pictorial Collections

Pictorial collections are used by scholars in a variety of fields: for example, architectural historians, environmental researchers, and social and political historians. While research libraries have focused primarily on scholars as their principal audience, the fact remains that scholars are no longer the majority audience of a significant number of historical pictorial collections. Such collections also are very heavily mined, directly or through professional picture researchers, by publishers, news magazines, advertising firms, and the like. Therefore, in addition to their use as primary documentary evidence, they serve to illustrate scholarly publications already written, "coffee table" picture books, educational electronic publications, news stories, and magazine advertisements. Users also acquire copies of items for personal reasons, both aesthetic and sentimental. As a result, libraries with pictorial collections increasingly are serving as stock photo agencies. The universal reach of digital images on the Internet provides a means to respond to this escalating demand to furnish broader access. Because separate pixels replace the continuous tone of the originals in digital images, artifactual value is lost, and the medium is capable of seamless manipulation. Digital images may, in fact, better address the needs of a non-scholarly audience.



BEST COPY AVAILABLE

Organization and Service of Collections

Historical pictorial collections are among the least accessible sources available to researchers because of their large size, complex organization, physical fragility, and often rudimentary description and cataloging. Most consist of large groups of related materials that share one or more significant common denominators, such as source, subject, or medium (e.g., the bequest of all 750,000 drawings, photographs, posters, models, etc. that constitute the work of Charles and Ray Eames; all 2,100 performing arts posters—theatrical, minstrel, and magic; all 600 daguerreotypes, etc.). That common feature often serves as the framework for organizing and providing access to the individual pieces.

These collections can be physically assembled and housed together as a single unit or organized intellectually under a single classification while being stored in separate groupings. Bibliographical controls can give primacy to the group or to individual items. They can be served as containers of material or one item at a time. However, for all of the variables, each collection tends to be characterized by two attributes: (1) a logical coherence that binds the contents together; and, as a result, (2) a totality that enhances the research value of each individual item beyond what it would have in isolation.

Within any one library, retrieval of pictorial items for use varies a great deal, reflecting the unfolding history of picture librarianship. Most are described much as personal archives are—that is, by a group record with one or more levels of general description within that. Staff provide users with groupings of images in one or more containers, with catalog cards and searchable electronic records giving succinct descriptions of groups of materials. In general, in the context of the reading room, only limited numbers of items are cataloged and provided to users individually.[5] Additionally, significant portions of these materials may be organized to make them available through self-service. Patrons can remove them from file cabinets or shelves in the reading room without the intervention of staff. These collections tend to be self-indexing, in that each artifact serves as its own record. Since each is filed and retrieved on the basis of annotations on its mounting, there is no catalog record, shelf-list, or finding aid for intellectual control.

When item-level information is provided for self-service or group-cataloged materials, it usually consists of captions and other annotations on the back or the mount of each individual image. The non-standard form of this data, which may be typed or handwritten, and haphazardly aligned, belies its importance. It can provide information about subject, date, creator, past ownership, and copyright restriction. Only with the item in hand can the researcher fully assess the data's veracity and significance. However, this rich mix of accumulated documentation challenges the capabilities of the digital environment. Even when images can readily be digitized, it is not al-

[5] As discussed below, this pattern often is reversed when collections are digitized and distributed electronically.

ways possible to capture the annotations efficiently because of the variety of their placement, alignment, media, and size. One choice is to scan the item in such a way as to capture all the annotated information, which challenges the capabilities of production-level scanning and could introduce an artificial "frame" to the image, adding separate scans of the back of the mount, or scanning at higher levels of resolution than are required for the images alone. The other choice is to provide the information in a database. Either alternative will introduce to any scanning project inefficiencies that will increase the resources that are needed and the time that is involved. Only those relatively small portions of historical pictorial collections that are individually cataloged and provide consistent item-level controls can be converted for use in the digital environment with a high degree of efficiency.

Preservation Issues

The variety of collections leads to numerous preservation concerns, which range from storage conditions to patron service. Some materials tend to degrade because of their chemical make-up (e.g., nitrate and diacetate negatives). Others have deteriorated because of past, and at one time acceptable, library practices (e.g., dry mounting photographs on acidic board). Watercolors and color transparencies are sensitive to light. Architectural drawings on tissue paper are inherently fragile and easily damaged by handling. Because of their size, large posters and other outsized materials cannot readily be handled by patrons. Materials such as glass-plate negatives are both extremely fragile and unique and are served to patrons, if at all, only under continuous curatorial oversight. Indeed, all of these historical materials are prone to damage from continual handling and inadvertent mishandling by researchers in the normal course of library business. Even if a researcher uses digital access for the very restricted purpose of determining which images he wishes to see in the original, the process will contribute to their preservation by reducing handling and exposure to light. This is one area in which digital images can greatly enhance core library functions.[6]

Digital Images and the Reading Room Experience

There are three conditions that provide a relatively consistent, and very real, context for most research that is conducted in a reading room using historical pictorial collections. The necessity of traveling to a reading room and the actions that are prerequisite to conducting research once there tend to limit the range of the user public. While in most cases this public will extend beyond the principal audience of research scholars, such factors as time, distance, potentially burdensome administrative procedures, various restrictions related to

[6] This form of conservation is quite distinct from the more controversial subject of using digital images as preservation surrogates for deteriorating photographic negatives, as discussed below.

the researcher's status (to include age), and collection security will keep reading room research from extending beyond serious users to the casual browser with any frequency. The physical reality of working in a specific reading room leads the researcher to a sense of the identity of the institution. Its facilities, staff, services, collections, quality, and seriousness of purpose soon become known entities. Finally, when a pictorial collection is served in a reading room, it provides a full range of information and cross-references to the researcher that establishes a frame of reference for any single image it contains. The images are there on the table, in their boxes and folders, as a cumulative presence with an insistent sense of the whole. They can be viewed individually or laid out in groupings, and these grouping can be revised, enlarged, and totally changed. Therefore, the researcher can bring evidence together that defines underlying patterns, establishes essential differences, or clarifies nuances.

The context established by doing research in a reading room provides a comfort zone for both the institution and the researcher in that it establishes parameters that govern expectations and assumptions. None of the conditions that establish this context pertain to gaining access to digital images on the Internet. No assumptions can be made about the probable audience; anyone can gain access to the images at any time for any purpose. The physical setting becomes everyone's computer at the office or at home, and the inter-institutional reach of the Internet obfuscates any consistent institutional identity. The sequential manner in which most digital images are viewed deprives the researcher of the flexibility that can be attained by grouping originals on a table and can result in individual images' being separated out and downloaded without the researcher's ever gaining a sense of the whole collection. Researchers can and do adjust to these differences from the reading room experience when viewing pictorial materials on the Internet. In many instances, the enormous advantage of being able to conduct research from remote locations more than outweighs the losses inherent in being far removed from a reading room.

Above all, viewing historical pictorial collections over the Internet cannot be a viable alternative to the reading room experience unless the integrity of the digital images can be assured. There must be the presumption that an honest effort has been made to replicate the original image digitally to the degree that the technological constraints allow. Any deliberate or unavoidable deviations must be documented (e.g., cropping or so-called lossy compression). In the enormous universe of pictorial material available on the Internet, this presumption of image integrity and documentation of the deviations cannot always be made. The situation is highly variable and includes a continuum of electronic publications that extends from the digital equivalent of scholarly journals to the tabloid press. All are equally available on the Internet. These publications will not necessarily be identified by distinctly different titles and physical features (e.g., layout), characteristics that clearly demarcate one type of publication from the other in the print world. Additionally, in terms of

the resources that are required, the process of "publishing" on the Internet is extremely accessible compared with print publication. A publication on the Internet can be created by a single individual pursuing a personal agenda beyond the reach of outside review.

Since a digital image reduces the tones and colors of the original to numbers, and the numbers can be manipulated, the digital images can be manipulated as well. Further, compared with the techniques that previously were available for manipulation of hard-copy images (e.g., cut and paste, air brush), the manipulation of digital images is more easily accomplished and the results can be virtually impossible to detect. Finally, in terms of both structure and physical appearance, there is no intrinsic difference between a digital image that is based on a genuine pictorial document and a digital image that has been manipulated—for example, that has been constructed from several pictorial documents. Therefore, digital images lack an essential characteristic of primary historical documentation. The physical properties of digital images provide no artifactual evidence of their authenticity that is equivalent to, for example, the evidence yielded in the originals by physical medium and support. Even if special techniques are employed (e.g., water marking and encryption), these can do no more than assure that there will be a continued association of the digital images with the institution that originally placed them on the Internet, even after they are downloaded and then made available on new websites. While these techniques provide the digital images themselves with a traceable provenance, they do not provide information about the relation of the digital images to the originals from which they were made. It is the reputation of the institution that first placed the images on the Internet that provides this linkage.

Doubts about the authenticity of any one digital image can be raised by apparent errors in the information it provides, and there is seldom readily available information for determining the accuracy or authenticity of the data found online. One of the uses of primary documentation is to provide new evidence that confirms or contradicts "facts" and conclusions that are widely accepted. In such cases, the document usually is subject to especially intense scrutiny in terms of both its physical makeup and the chronology of its provenance. Such tests cannot be applied to pictorial documentation in digital form. In this context, the source of the images and the documentation of their technical specifications take on a special importance. Iastitutions must provide complete information about themselves, the collections being digitized, and the technical specifications of the digital images, as well as use such techniques as encryption and water marking. On arriving at any site on the Internet, researchers should assess this information with intense critical judgment.

II: Observations about Images in a Digital Environment

"Look at the pictures," John Vlach writes about viewing the HABS documentation of slave quarters. "Pore over the drawings. Check the details. Do it carefully, and you can develop almost a tangible sense of the buildings that once sheltered the everyday routine of slaves."[7] What Vlach here alludes to is the most beguiling quality of an original document—the effect that its authenticity as an artifact has on the beholder. The artifactual nature of a given document is an irreducible feature, one that is lost in any type of reproduction or reformatting, be it in print, on the computer screen, or on a slide. In speaking of digital access to historical images, one must acknowledge, as with any surrogate, that digitization cannot serve as a replacement for the original.[8] The issue, rather, is how well it serves as a simulation of the original, and what properties it might have that make research easier. These attributes of the digital image will be explored here, so that we may better understand the advantages and disadvantages of digitizing historical pictorial materials.

> *Digitization of information is the single fundamental theme underlying modern data communications networks and their usefulness. Images, speech, music, diagrams, and the written word can all be translated into a sequence of numbers. That sequence can be stored, processed, and/or transmitted. At the other end of the process, the numbers can be interpreted to recreate information in the form in which it was originally expressed. Since computers are able to manipulate digitized information, advances in computer technology are now applicable to information of all types, and all types of information can be transported on data communication networks.* [9]

Access to historical pictorial collections through digital images can be quite beneficial to scholars and other researchers, especially if the images are linked to a searchable database. The quality of digital images can be sufficiently high to serve a number of research requirements, sometimes obviating the need to view the originals. At a minimum, one can create image quality that is sufficient to allow a researcher to narrow down the number of items that must be called from storage to be seen in the original. In almost all of its aspects, ac-

[7] John Vlach, Preface, p.xiii.

[8] Conversely, digital images, when published on a CD-ROM, are an excellent replacement for illustrations in print publications. The CD-ROM can contain many more illustrations; they avoid most of the problems inherent in gaining access to images over the Internet, and they are less subject to image degradation. On this latter point, see Michael Ester, *Digital Image Collections: Issues and Practices*, Commission on Preservation and Access, December 1996, p. 6.

[9] Douglas E. Van Houweling and Michael J. McGill, *The Evolving National Information Network: Background and Challenges*. A Report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access, July 1993, p.3.

cess to digital images offers improvements over accessing original pictorial materials in a reading room without the added benefits of digital images. The search is easier, access is faster, downloaded records and thumbnail images are available, and, as a bonus, handling of the originals is reduced, which contributes to their preservation.

The Digital Infrastructure

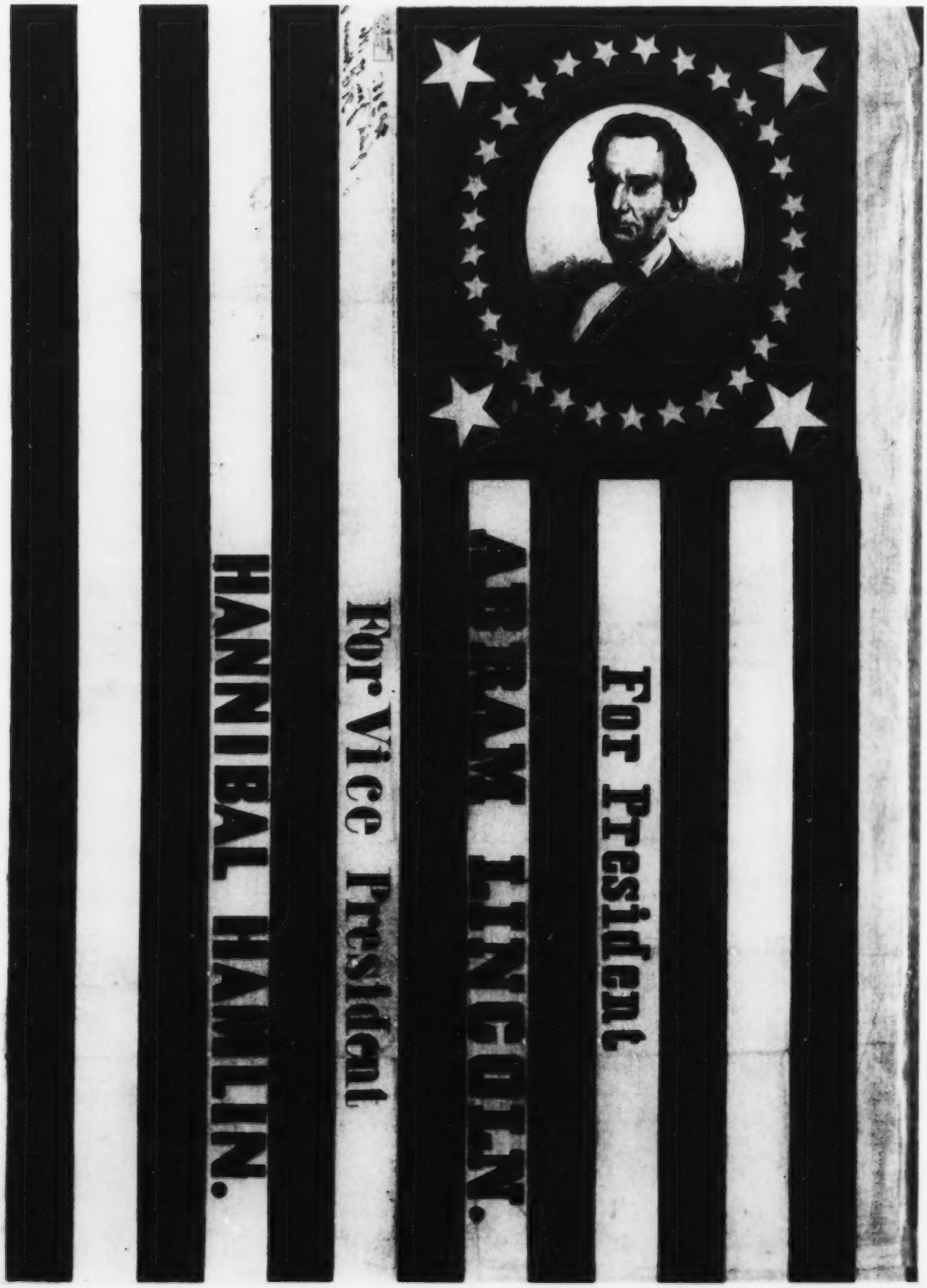
Currently, most local area networks and the Internet itself cannot easily and quickly present high-level resolution images and may thus limit rapid access to the details that they contain. [10] At the present stage of technology, higher levels of resolution require inordinate amounts of time for the information to be transmitted from the file server to the work station, and, even then, the image may only be viewed on the monitor a segment at a time. While compression of digital files may improve speed of delivery, the most efficient compression techniques currently available for pictorial images contribute to image degradation, the very problem that engendered the need to digitize at higher resolutions in the first place. Creation of multiple versions of digital images, so that they are available at various resolutions, allows researchers to select the format most suitable for their purposes. It also recognizes the diverse capabilities of the equipment, software, and transmission lines being employed by those who are not served onsite. This is an important point for those institutions committed to making their images as widely accessible as possible.

Additionally, unless each part of the digital system has been calibrated, there will be deviations in image quality at each stage from capture to display on a monitor or output on a printer. Even assuming that capture and processing have been calibrated by the institution that created the digital images, images will vary from monitor to monitor no matter what the capability of the monitor in terms of spatial resolution unless their tonal resolution is calibrated to the specifications of the digital file. In terms of image quality, the differences could be significant for research and scholarship.

Purpose, Audience, and Resolution

The degree to which digital access is a satisfactory alternative or an improvement over using the originals also depends on the visual characteristics of the originals and on the nature of the research being conducted, which have nothing to do with digitizing *per se*. A political poster, with its large forms and broad areas of color, was meant to be seen and readily comprehended at a distance. Thus, a

[10] Such high-resolution digital files contain enormous amounts of information. For example, the file for an 8 x 10 inch, continuous-tone, black-and-white photograph would contain as many bits of information as a two-million-word essay. (This example assumes scanning at a spatial resolution of 4000 x 5000 pixels and 8 bit per pixel gray-scale tonal resolution; and words in the essay averaging a long ten letters each.)


For President
ABRAM LINCOLN.
For Vice President
HANNIBAL HAMLIN.

high-resolution digital image of a banner for Lincoln's 1860 presidential campaign is liable to prove a more satisfactory surrogate for a scholar than will a digital image of a Civil War photograph of officers on the field (**Figs. 6,7**). Yet, the researcher who uses the photograph as the illustration to put on the cover of a brochure most likely

↑
[Fig. 7] Alexander Gardner, *Lieutenants Wright and John W. Ford of Averil's Cavalry. Westover Landing, Virginia.* Wet collodion glass negative. August, 1862. Civil War Photograph Collection. (LC-B8171-0622)

←
[Fig. 6] H. C. Howard, printers, *For President ABRAM LINCOLN. For Vice President HANNIBAL HAMLIN.* Woodcut or lithograph printed in red and blue on cotton. 1860. (LC-USZ62-14844)

will find the digital surrogates satisfactory, while the historian whose research resides in the details of uniforms, insignias, and field equipment probably will not. However, even the Civil War historian might find digital surrogates satisfactory as a way of identifying which of the photographs need to be seen in the original, thereby adding new efficiencies to the search. At this juncture, basic tenets of librarianship and digital technology are inexorably joined.

The implications of these observations are critical to the practical decisions about which collections should be digitized and which levels of resolution should be used. Any decision about resolution also is a decision about the minimum level of detail that must be readable to satisfy the uses to which the images will be put. Critically important questions follow immediately: who are the target audiences, and

BEST COPY AVAILABLE

how will they most likely use digital images of the collection in question in their research strategy?

Decision-making about optimum image resolution should begin with an analysis of how digital access will address the needs of scholars and other audiences. Further, this decision should be collection-specific. In the examples given above, digitizing posters at a moderately high but nevertheless practical level of resolution probably would satisfy most of the needs of all audiences, who in all likelihood could conduct research without having to consult the original artifacts. This would not be so with architectural photographs. Yet, in all likelihood, digitizing them at the highest levels of resolution would neither obviate the architectural historian's need to examine the originals nor significantly improve the usefulness of the digital images as a means of efficiently identifying which of the images had to be viewed in the original. Nor would higher resolution make these particular digital images significantly more useful to other audiences.

There are, in fact, many repercussions to striving for a much higher level of resolution than is absolutely necessary for the specific situation. Images digitized at higher levels of resolution contain increased amounts of information, leading in turn to increased costs to digitize, longer transmission times, the necessity of more storage space on file servers, and the need for upgraded researcher workstations, with more memory, and higher spatial resolution of the monitors, to gain access to the higher resolution images at a practical level of efficiency. Each decision about resolution, in other words, affects the costs of a project, the amount of staff time that must be devoted to it, the hardware and software needed for access, and, of course, the capacities of the information technology infrastructure of an institution.

The Pace of Change

No discussion of these issues would be complete without reference to the pace of change in digital technology. This is pointedly stated in a 1993 report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access.

> *All the evidence is that there will be a continuing technology explosion in all these fields—communications, processing power and storage—for at least two decades to come. If one can imagine it and if it will be useful, the fact that an application is not currently feasible is only a temporal statement and quite likely to be overcome by technology in the near future.*[11]

In light of the history of the past decade, this statement of six years ago applies to the creation of digital images as well. For example, an initiative for direct satellite transmission of information from

[11] Van Houweling and McGill, pp.1,2

server to workstation in digital form would allow detailed, very high resolution images to be transferred efficiently to researchers.[12] The fast pace of technological development must somehow be taken into consideration when planning a digitization project, which presumably reaches out toward a longer future.

III: Digitizing Historical Pictorial Collections

Why Create Digital Images?

Decisions about resolution are usually among the last to be made, well after the basic decisions about selection criteria. Among the reasons for digitizing historical pictorial collections are the following: to improve access, to preserve original materials, and, in the case of some repositories, to extend institutional outreach and public relations.

We have identified certain features of archival holdings of pictorial materials that are integral to their value for researchers: the completeness of collections; their contiguity with other collections of like or complementary materials; their being authentically historical; and, of course, their being original items, not simulations. All but the latter quality can and should be captured to the highest possible degree in a digital environment, at least as long as the audience is the community of researchers and scholars that these repositories have traditionally served.[13]

When comparing digital access with the reading room experience, it must be remembered that incompleteness is the underlying reality of universal accessibility. Regrettably, institutions with large pictorial holdings probably will never be able to digitize more than a modest portion of the collections in their custody. At least at present, digitizing is an incremental process that involves choices as to which collections (or, equally regrettably, portions of collections) will be digitized and in what order.[14] Within this limitation, maximum ad-

[12] Plans for this system, *Teledesic*, call for it to be operational by 2002. For details, see the Web page at http://www.teledesic.com.

[13] Since digital images do not contain artifactual evidence, the historic authenticity of the originals is derived by the researcher from the documentation that is provided and the reputation of the institution that places the images on the Internet.

[14] There is no comprehensive inventory of all library programs that are creating electronic versions of their historical pictorial materials, and no statistical analysis of the number of digital images in relation to the number of total pictorial holdings. Some sense of the minimal reach of digitizing initiatives when compared with the cumulative corpus can be garnered from the *Digital Collections Inventory Report* by Patricia A. McClung (February 1996, sponsored by the Commission on Preservation and Access and the Council on Library Resources; also available at http://www.clir.org/cpa/reports/mcclung). It also is instructive to note that the National Digital Library Program's stated goal of making 5 million items available over the Internet represents less than 5% of the Library of Congress's holdings.

vantage can be taken of the new capability if interrelations among collections are given a very high priority when deciding which will be digitized. As a result, at each incremental step, the aggregate of collections within any defined subject area that are available digitally, while not comprehensive, will remain complementary, and their research and reference potential will be maximized. This strategy becomes all the more effective if the selection is coordinated with other institutions that have related collections.

The limitation of digital access to only a fraction of the collections that would be available in any one reading room is unavoidable. This is quite different from limiting the span of access within those collections that are digitized. In terms of text-based holdings, this would be the equivalent of making monographs available electronically on the basis of a selective bibliography rather than on the basis of the library's catalog records of all of its holdings within a given subject area. In both cases, researchers would find that the advantage of increased ease of electronic access was countered by the disadvantage of limits on the span of access that would not be encountered if these materials were consulted onsite. Heretofore, researchers in these large collections have been able to rely on the comprehensiveness of an institution's holdings as one of the innate qualities of the resource. To limit digital access within a collection only to selected images, rather than to the entire corpus, undermines the documentary function of historical pictorial collections, whose cardinal strength derives from their completeness. In order for digital surrogates to serve traditional purposes of primary source research, users must be able to search comprehensively without being limited by editorial choices that are guided by other purposes. The danger is that one may use a subset of readily available selected images as if it were a perfect subset of the whole, thereby eliminating from consideration the remaining images and the potentially different documentation that they might provide. Scholars will still need to consult the full record to understand accurately the nature of the source base.

Remote digital access to selections from a collection can help prepare scholars for subsequent research, especially when the principles of selection for digitization are made clear. If selections are accompanied by a database documenting the scope of the complete collection, the scholar can assess what part of a collection the digital images represent and gain an understanding of the nature and quality of the images from the sampling. Even under these highly controlled conditions, careful selection of items in a digital collection primarily targeted at sophisticated scholars remains a challenge.

Primary and Secondary Education

Editorial selection within collections for primary and secondary education is quite a different matter from primary source research, and can add outstanding value for educators. The types of images that are found in historical collections are rarely available through the

schools for students in primary-, secondary-, and even college-level courses. The opportunity to expose young students to the riches of primary visual resources is one of the most exciting features of the digital era. Research libraries must decide what priority should be given to this extended audience, especially if it comes at a cost of service to scholars and researchers.

The weighing of priorities must come very early in the process of selection. While it is physically possible to scan and describe a selection of items, and then go back to that collection at some later date to "finish the rest" or add items, it is quite unrealistic to expect most institutions to do this routinely. Given the scope and complexity of digital conversion projects, the amount of staff time involved, the physical handling required, and the efficiencies of working with a collection as a whole (or a coherent portion of the collection), an institution cannot afford to expend its resources by returning to a project once it is over. Further, good archival practice mandates that materials be handled a minimum number of times. If one is going back to a collection time and again, and sorting through it while selecting additional images to digitize, something has gone very wrong in planning.

However, if a research institution has made it a priority to reach a new educational audience—say, K-12—it is possible to make selections from digital files of whole collections to create a subset of images and records useful for those levels. There certainly are cases in which an institution that has traditionally served only scholars on site might wish to expand access to a younger group. The Library of Congress and the National Archives, both with rich, unique historical resources and both funded by the taxpayer, have made a concerted effort lately to reach out to a new national audience through Internet and CD-ROM access. Other research libraries might best focus on digitizing entire collections, while developing partnerships to create the subsets that best serve primary and secondary education. Commercial companies have long mined historical pictorial collections to create illustrated books and sets of slides for educational purposes, and grant-giving organizations are now demonstrating a growing interest in the educational possibilities inherent in digital technology. Such partnerships would allow research libraries to exploit more fully the educational potential in digital files of whole collections, gain access to expertise on curriculum needs that usually is not at their disposal, and quite possibly supplement their resources while so doing.

Retrieval and Service

A good searchable database undergirding a collection of digitized images can improve access by making the search for particular items more efficient and by bringing lower resolution or compressed, higher-resolution images swiftly onto the screen. In some circumstances, depending on the organization of the digital files and their records, this searching can be done across collections. Under current condi-

tions, using digital images as *reference surrogates* may provide the most significant contribution to improving research procedures. The images can be enormously useful in the earlier stages of the research strategy, both remotely through the Internet and in the reading room as well, in that they allow researchers to pinpoint the specific items in a collection that they would want to study in the original or in higher-resolution copies.

Image files, when used as reference surrogates, often are accompanied by high-resolution uncompressed digital images. These images can serve as *research surrogates* that are used at the final stages of the research strategy. Depending on both the nature of the original image and of the research, these high-resolution images might serve as a substitute for seeing the original (or for ordering a photographic copy, if the image is to be reproduced). This would be the case, for example, with the political poster and Civil War photograph discussed earlier. **(Figs. 6,7)** However, in most cases, research surrogates tend to be more satisfactory substitutes for the original as their use moves away from that of primary historical documentation. The high-resolution digital image of the Civil War portrait photograph is far more likely to satisfy the need for an illustration of a history of the war than for studying the insignias and field equipment in detail.

As visual collections become more vulnerable to damage, and as their monetary value and susceptibility to theft increase, the current trend toward more restrictive access to the originals will accelerate. The use of digital images as research surrogates will grow accordingly, with access to the originals restricted to those circumstances where it is absolutely necessary. The Library of Congress's Farm Security Administration (FSA) photographic collection is a case in point. As part of an initiative that includes copying the deteriorating nitrate and diacetate negatives, all of the images in the collection are being digitized. In keeping with current P&P policy, when the research surrogates of these images become available, they will in all probability be served in lieu of the originals, unless viewing the original is absolutely necessary to the research.

Digital- and Film-Based Preservation Surrogates

Although the long-term persistence of digital files *per se* is uncertain, institutions are investing heavily in this medium, relying on data migration to assure the files' continued existence.[15] Even if this guaranteed that they would last a very long time—longer than film-based preservation media—they would not, at present, be the preferred medium for preservation of pictorial materials because of the image degradation that occurs during conversion. Nevertheless, it can be argued that, under specific circumstances, serious consideration should be given to using very high resolution digital images as *preservation surrogates* for certain types of continuous tone, black-and-

[15] See *Preserving Digital Information, Report of the Task Force on Archiving Digital Information*, for the Commission on Preservation and Access and The Research Libraries Group, Inc., May 1, 1996.

white photographic negatives. Such a strategy would take advantage of the degree to which digital images approximate film negatives as a medium, the current practice of creating digital images in addition to preservation film copies in order to provide access to negative collections, and the lower costs of creating digital images rather than high-quality, film-based preservation copies.

Digital preservation surrogates require tight quality control, complete documentation of the technical specifications, and calibration of all parts of the digital system, from capture to image display or printed output. Constant calibration can be achieved and maintained when these high-resolution images are used within an institution, where all of the elements can be controlled. The quality control and documentation are necessary because the assumption must be that the surrogates will outlast the originals, which will continue to deteriorate and eventually no longer be available to verify the degree to which they are accurately reproduced in the digital images. Complete documentation of the technical specifications is required because it is necessary to approximate the original images as closely as possible in the first place and to verify this relationship over the long term, even after the digital image has been subject to migration a number of times.[16]

The option of using digital images as preservation surrogates can be a valuable supplement to (though certainly not a substitution for) an ongoing preservation program that follows such traditional strategies as improving storage conditions (with special reference to temperature and humidity control) and using film as a medium for preservation copying. Ideally, this option should be considered only if preservation copying using film is not possible and the choice is between using digital images as preservation surrogates and doing nothing at all. There are enormous numbers of photographic images, for the most part negatives, deteriorating because of their chemical make-up. Few institutions have sufficient resources and time to copy all such negatives before some of them deteriorate to a point where they no longer are useful. In such cases, the choice is between using digital images as preservation surrogates for certain of these images and losing them. Though more image degradation occurs when making preservation copies by digitizing directly than when copying onto film, does this difference in degree have an equally overriding importance for all collections? The current method of making film preservation copies of deteriorating negatives also involves image degradation as it moves successively through several generations from the original negative, to the interpositive, to the preservation negative, to the copy print.

If it is admitted that there are specific circumstances under which high-resolution digital images might be used as preservation surrogates, what defines the characteristics of the photographic negatives that might be considered for preservation in this manner? Michael

[16] Michael Ester's essay provides useful insights (*op. cit.*). See especially his discussions of reproducing original objects (with its insistence on color and gray-scale control bars), assessing image quality, and documentation (pp. 7-12, 17-19).

Ester provides a starting point for the definition in his discussion of archival digital images (i.e., digital *preservation surrogates*). Arguing that it should be possible "to use the archival digital images in any of the contexts in which the photographs...would be used," he goes on to suggest that "photographic reproductions" should be the "archival standard," going "film to film."[17] Extrapolating from his arguments and applying them to the situation of deteriorating film negatives, one could make a parallel comparison between a copy print that is made from a preservation negative and a copy print that is made from a digital preservation surrogate. In a fundamental departure from current practice, the final products that are delivered to the researcher—the copy prints—would become the points of comparison.[18] Decisions would be based on the different degrees to which film-based and digitally based copy prints would satisfy the requirements of the researcher for each of the contexts in which they would be used.

The number of variables and subjective judgments involved in each decision prevent establishing any systematic hierarchical basis for when digitally based preservation of deteriorating negatives might be considered a viable alternative to film-based preservation. Rather, one should analyze the relevant factors on a case-by-case basis. Principal among the variables are the nature of the original negatives in terms of their internal aspects (e.g., physical characteristics and subject content); the nature of the collection from which they come (e.g., the original purpose for which they were made, the historical importance of the collection, the nature and historical importance of the creators); and the contexts in which the researchers probably will use the copy prints. A conservative approach would favor film-based preservation when a collection has but one significant factor that mitigates against digitally based preservation. However, such a conservative approach will not be particularly helpful to an overall preservation program, in which judgments cannot be postponed indefinitely without drastic consequences.

Collections at the extremes are not the problem. A collection of images that would no longer be useful as documents because too much detail would be lost in the digital files (as is the case with many architectural photographs), or that in the aggregate are a monument in the history of documentary photography (e.g., the Farm Security Administration), or that were created by a historically important photographer (e.g., Ansel Adams's *Manzanar* series) should be preserved on a film-based medium. Conversely, a collection of images that are largely dependent on overall effect rather than particular details (as is the case with much photojournalism),[19] or that

[17] *Ibid.*, pp. 11-12.

[18] In film-based preservation projects, quality control is conducted by comparing first generation interpositives with the original negatives.

[19] Even such morgues might require film-based preservation for certain groups of negatives when one of the relevant factors applies; for example, John Vachon's images in the LOOK Magazine collection are the work of a "historically important creator."

have no particular historical or aesthetic importance in the aggregate or because of their creators, might well be a candidate for digitally based conservation if the alternative were probable deterioration.

Between these extremes are great numbers of images about which decisions are not as clear-cut. Continued improvements in digital technology, coupled with the inevitability of growing pressures to accelerate preservation copying of negatives as they continue to deteriorate, will first force the issue in reference to the more obvious examples, and then begin to encroach on this vast middle ground.

IV: Beyond Images: Rights and Records

Digitization projects are major initiatives that consume large staff and financial resources and can continue to do so long after the initial projects have been completed. They are labor-intensive, require extensive expertise, are organizationally and logistically complex, and demand a long-term commitment to continued maintenance, migration, and updating. These operations extend far beyond the relatively straightforward task of digital capture and transmission, the focus of this discussion thus far.

Some sense of the seriousness of the required commitment can be seen in the *Project Planning Checklist* published by the Library of Congress's National Digital Library Program (Appendix A). Even though written in outline form, the checklist runs to more than three pages of activities. Although many of the tasks it enumerates take place well before, or after, the core process of producing the digital files, they are no less critical to the success of the project as a whole. These tasks often involve extensive analysis, planning, and research—activities that tend to be labor-intensive and potentially expansive.

Most of the tasks in the *Checklist* are related to the essential nature of the images under consideration. Two areas, however, the issues of rights and of records, are separate topics from image presentation, and of special import for institutions. Together, they bear directly on making the image collection available and accessible to the general public over the Internet.

Rights

Since distribution of images over the Internet is a form of publication, the process is subject to the laws related to rights (copyright, the right of privacy, and the right of publicity) and to any restrictions that might have been placed on a collection by the donor. Failure to take heed of the laws governing these rights can lead to potentially significant financial penalties. A good introduction to the laws, and to some of the more accessible circulars, handbooks, and guides on

the subject, is *Copyright and Other Restrictions Which Apply to Publication and Other Forms of Distribution of Images: Sources for Information*, a Reference Aid published by the Library of Congress's Prints and Photographs Division (Appendix B). The Reference Aid reinforces the cautionary warning provided by the *Checklist*. From the ambiguities of fair use when applied to the digital environment, through the complexities of the duration of copyright protection, to the nuances of right of privacy, the Reference Aid speaks to a need for informed, professional advice and extensive research, applied on a case-by-case basis.

A number of strategies are available to comply with copyright law, or to avoid the issue altogether. For example, digitizing and distribution can be limited to those collections that are clearly without copyright restriction because of their age, or in the public domain as a result of their source (e.g., nineteenth century photographs, or those created in-house by Federal agencies). Otherwise, permission must be sought from the copyright holders, who can be as few as one for the whole collection and as many as one—or more—for each item in the collection.

Gaining permissions to digitize, like gaining permissions for any type of in-house publishing project, can involve massive amounts of staff time and often costly expert legal advice as well. The research involved in identifying and locating the copyright holder(s) can be labyrinthine, in part because of the difficulty of identifying all the rights holders. An institutional donor of a collection may have owned only the physical, not the intellectual, property for many of the images. Records may not indicate which institutional staff created the images, or, if contractors created the images, whether they maintained copyright or worked for hire. Copyright may have passed on from the original creator to heirs who are not identified in the available documentation. Once the copyright holders are identified by name, an equally arduous search may be necessary to locate them geographically before negotiations for using the images can begin. The final agreement will probably involve a payment of fees, though that might be reduced or eliminated by addressing concerns about commercial exploitation. For example, limitations could be placed on the distribution of the digital images by making them available only at specific sites (e.g., educational institutions), or by providing only low-resolution thumbnail images that, when downloaded, do not allow satisfactory reproductions. Other alternatives include charging end-users when they download the images, which in turn requires electronic tracking and payment, and imbedding an identifier, such as a watermark, in the images.

These strategies will prove satisfactory only if the restrictions do not compromise the purposes that the digital images are meant to serve. Programs to digitize historical pictorial collections are most effective when they make multiple, thematically related collections universally accessible on the Internet as both reference and research surrogates. Limiting digitization to those collections that are without copyright restrictions will significantly reduce the scope of such pro-

grams; site licensing limits their reach; and providing thumbnail images alone limits their usefulness to researchers. Further, any one of these undertakings—research, site licensing, negotiating agreements, fees—will have a significant impact on both the costs and schedule of the digitization initiative. Finally, equally burdensome activities also may be necessary to address rights of privacy and publicity.

Records

A rich mix of information should be made available to researchers along with the digital images. Some of the information can be provided on a collection level. The identity and nature of the initiating institution provides credentials for the images, and technical specifications indicate the degree to which they can replicate the originals. A description of the collection as a whole, e.g., provenance, size, and chronological span, provides a context for each of the images it contains. Additionally, each and every image must be linked to an access record, and these can range from item-level records using the MARC format, through transcribed captions searchable by key words, to collection-level finding aids, with SGML encoding, providing linkage to individual images or groups of images.

Ideally, an extensive amount of information should be provided for each image: title, medium, dimensions, creator, restrictions, and the call number, reproduction number, and frame identification. In addition, the image should be searchable by subject, through subject headings, key words, or the image's organizational location in the finding aid. Linked item-level records can provide the most information and are the most searchable, both within and among collections. However, this linking may prove impractical for sustained digitizing programs of collections of any significant size, as it is quite labor-intensive to create. Encoded finding aids are the most efficient to create in terms of both time and costs. Extensive work is being carried out in many forums on the problems of documenting digital images.

As discussed earlier, most historical collections are group-cataloged, if they are cataloged at all, and, traditionally, their finding aids are usually more minimal by far than, for example, those for manuscript collections. Much of the information needed for access records appears as annotations on the mounts of the original images. Therefore, no matter what the form of the access records, information will have to be assembled from various sources. In brief, because of the inherent characteristics of most historical pictorial collections, access records for digital surrogates are not ready-to-hand and involve far more than simple digital conversion of records. Rather, such records most often must first be created as part of the digitization project, which contributes to the scope, cost, and duration of the project.

V: What to Digitize: The Questions to Ask

Decisions concerning which historical pictorial collections should be digitized, and in what order, are best approached step by step, process by process. The sum-total of information that is gathered will provide a strong factual basis for making the key decisions, establishing the schedules, and determining the needed resources.

The Intended Audiences

Whom have the institutions defined as their audiences, scholarly and popular? What is the audiences' relative importance in terms of the institutional mission? For what scholarly or popular purposes does each audience use historical pictorial collections? How will the digital images serve these purposes (e.g., as reference and/or research surrogates)? What attributes of the digital images are most important for each of these audiences (e.g., fidelity to the original, speed of access)? Overall, who will be served by digitization and how well?

The Collection as a Whole: Subject Content

What broad subject is documented by the collection? Will it be of interest to the institution's various audiences (i.e., if made available over the Internet, will it be used and how widely)? Is the collection coherent (i.e., do the images inform each other)? Is the collection related to other collections that will be (or have been) digitized? Is the relationship redundant or complementary? Does the collection contain coherent subsets of images that could lead to cooperative agreements with commercial partners because of their appeal to special audiences (e.g., K-12)?

The Collection as a Whole: Physical Attributes

What is the size of the collection? Does it require physical processing or preservation before it can be digitized? Is the collection relatively uniform or diverse in terms of the types, sizes, and media? Is it easily handled (e.g., what is its fragility or size)? Does it have intrusive housing (e.g., overlapping mats, Mylar sleeves)? In light of its physical attributes, does the collection readily lend itself to digitizing, and, if so, how extensive will the digitization project be?

The Individual Images

Generally, what is depicted and in what way? Does the nature of the original image minimize the effect of image degradation (e.g., posters)? Conversely, to what degree is the usefulness of the images reliant on the legibility of details (e.g., architectural photographs, line drawings, and integrated text)? Does this reliance on detail preclude

the digital image's serving adequately as a reference or research surrogate?

Supplementary Benefits

Are there special circumstances involving other collections that would result in additional benefits accruing to the digitization project? Are items especially vulnerable to theft (e.g., items that are small and of high value)? If the collection is made up of deteriorating negatives, to what degree does it meet the recommended criteria for considering the use of digitally based, rather than film-based, preservation surrogates?

Rights and Restrictions

What is the status of the documentation about rights and restrictions? Is the collection protected by copyright? Does the subject content of the collection create the potential for privacy and/or publicity rights' being applicable? Are there donor or other restrictions? To what degree might these preclude or limit the distribution of the digital images beyond the reading room? What complexities are involved in seeking the necessary permissions (e.g., are there numerous copyright holders)? What are the potential costs of gaining those permissions?

Records

How do researchers gain access to the collection in the reading room? If there are bibliographic records for the collection, are they dependable? Are they group- or item-level records? If group-level, are they extensive or limited? Are there supplemental finding aids? If not, is the collection organized in a manner that would lend itself to using a finding aid as a means of access? Is the information needed to create records in a central location or dispersed? If dispersed, is significant information, not found elsewhere, appended to the images or mounts? To what degree can information be provided effectively on a collection level (e.g., information about creator, rights, medium)?

Although the questions are grouped under different headings, many are inexorably intertwined. They may not be comprehensive or applicable to all situations, but they do provide a starting point for analyzing the viability of a proposed digitization project.

VI: Conclusion

Taken together, all the efforts that go into digital reformatting programs, from planning and finding funding to long-term rights management and data refreshment and migration, should be put in the context of how such programs accord with and advance the mission of an individual institution. For some research libraries, defining that mission in terms of service to scholars is no longer axiomatic. Even as the potential of digital technology makes consensus about mission all the more essential, it is making such consensus all the more difficult to achieve.

Digital technology is moving libraries in dramatic new directions, and may be said to be not evolutionary but revolutionary. In the process of digital conversion, these images lose essential attributes as primary historical documentation, so fundamental to the traditional scholarly uses, while they are placed at the ready disposal of a vast audience of alternative users. In that sense, digital conversion projects differ significantly from preservation microform reformatting projects, which produce surrogates (film or fiche) that are also used primarily by scholars and other serious researchers.

The promise of the "National Digital Library Program—A Library for All Americans," cited at the beginning of the report, can be considered both auspicious and ominous. The assertion that wider electronic access to our national resources, including the visual record of our past, can enhance education, promote knowledge of our heritage, make America more competitive in world markets, and nurture an informed citizenry is a very ambitious and idealistic premise on which to base digital collection development. At the same time, it places a heavy burden on a young technology, challenges long-held assumptions about the scholarly mission of most research libraries, and suggests strategies that may shift resources from traditional core activities.

Only time will tell which parts of that vision become real, and to what degree they may alter the very nature of research libraries in so doing. Each institution wanting to explore for itself what digital conversion can do for its collections, staff, and patrons must carefully balance the revolutionary potential of such initiatives with traditional values and core services. This should begin with thoughtful deliberations about the institutional mission, careful assessment of which projects best serve its defined goals, and detailed planning.

Appendix A

Note: The following is from the Library of Congress's website at http://lcweb2.loc.gov/ammem/prjplan.html

Library of Congress, National Digital Library Program

NDLP Project Planning Checklist

This document outlines the production process for historical collections at the Library of Congress and reflects that institution's administrative structure and procedures. Not every collection requires all of the steps listed; some collections require additional steps not listed. In practice, many of the operations are carried out in parallel and not sequentially.

I. Select a collection for digital conversion

- A. Analyze collection
  1. Determine scope or extent of digitization (entire or subset?)
  2. Assess status of custodial division processing and housing
  3. Assess the status of access aids (degrees of completion, readiness, and format)
  4. Assess best format, e.g., full text conversion, scanned page images
  5. Assess the physical condition and readiness for scanning
  6. Assess restrictions and copyright
- B. Consensus on collection among custodial div., NDLP team, and Library admin.

II. Plan the approach to digitization

- A. Develop method and resource plans for collection preparation and digitization
  1. Develop plan for required processing by custodial division
  2. Develop preservation treatment plan
  3. Complete evaluation of physical condition with recommendations
  4. Determine formats for capture, archiving and presentation
  5. Determine physical size (number of characters, images) and special production requirements
- B. Determine repository requirements
  1. Determine scheme for file name assignment
  2. Register aggregate name for collection
  3. Estimate required storage space for digital collection
  4. Update NDL forecast for storage
  5. Evaluate existing finding aids or bib records and develop plan for access aid
  6. Develop plan for framework
  7. Develop restriction plan and implementation (copyright, terms of gift, publicity, and privacy)
     - a) Find and record restriction facts at collection level
       - (1) Search collection files, copyright records, exchange and gift records
       - (2) Create a narrative "findings" statement
     - b) Find and record restriction facts at the item level
       - (1) Search item files, copyright records, exchange and gift records
       - (2) Create fielded/tagged note in access aid

BEST COPY AVAILABLE

c) Draft proposal for actions to be taken prior to and at the "release" time
   (1) Pre-release seek required permissions
   (2) Pre-release: add notices to all restricted items
   (3) Release-time: provide only local or licensed-site access for restricted items
   (4) Release-time: provide no access to items restricted until a given date
d) Draft restriction statement to accompany online collection
e) Review copyright restrictions
   (1) Review facts
   (2) Draft proposal for actions
   (3) Draft restriction statements with advisors in Copyright Office
   (4) Forward findings, action proposal, and draft restriction statements to General Counsel for approval
   (5) Revise action proposal and restriction statements after General Counsel review
f) Implement action plans
   (1) Implement pre-release actions, e. g. , seek and obtain permissions
   (2) Implement release-time actions; provide access to collection

8. Workplan for digitization and access aid completed

III. Produce digital collection and access aid

A. Process and house collection
B. Implement preservation treatment plan
C. Item capture
   1. Preparation
      a) Prepare targets
      b) Prepare scanning instructions specific to collection
   2. Image capture
      a) Scan collection
      b) Process scanned images
      c) Review images for quality
      d) Coordinate rework
      e) Notify contractor of acceptance of images
   3. Archive images in repository
   4. Text capture
      a) Prepare keying instructions specific to collection or batch
      b) Mark up and key text
      c) Review completed text for quality
      d) Coordinate rework
      e) Process text into final form
   5. Archive text in repository
   6. Audio capture
      a) Create preservation and working copy
      b) Determine sample rate and resolution

c) Select digital audio file format
d) Analyze storage requirements
e) Acquire temporary storage space
f) Perform analog to digital conversion
g) Edit digital files removing "dead air" at cue-up points
h) Perform quality review
(1) Inspect graphic waveforms for truncation and peak
(2) Audition percentage of sound files

7. Video capture

D. Access aid development
1. Modify existing finding aid
a) Develop keying instructions
b) Photocopy and mark up existing printed aid
c) Coordinate off-site keying
d) Mark up according to EAD
e) Review finding aid for accuracy and completeness
2. Create new finding aid
a) Verify final processing and arrangement of collection
b) Draft finding aid
c) Mark up according to EAD
d) Review finding aid for accuracy and completeness
3. Item-level finding aid (Bib record-style)
a) Set up Minaret database and import item-level records
b) Upgrade preMARC or MARC records
c) Download available preMARC or MARC records
d) Add supplementary notes into Minaret records
4. Incorporate basic-level links
5. Add enhanced-access links or subject terms
6. Prepare collection-level MARC record for future inclusion in MUMS

E. Access aid complete

IV. Store in digital archive

A. Store files in directories as specified by naming scheme
B. Register items in URN handle-server (when in use)
C. Deposit items in digital repository (when in use)
D. All items stored

V. Create framework

A. Draft framework components
B. Review completed framework components
C. Create mockup of HTML document
D. HTML mockup approved
E. Develop and insert hypertext links

F. Coordinate search engine link with ITS
G. Insert final links
H. Add graphic enhancements to HTML pages
I. Mount HTML pages on LCWEB server
J. Review framework for accuracy and completeness
K. Framework completed

VI. Assemble digital collection

A. Store access aids in directories as specified by naming scheme
B. Register document-style access aids in URN handle-server (when in use)
C. Deposit document-style access aids in digital repository (when in use)
D. Generate indexes for related MARC records
E. Generate indexes for textual items in collection
F. Prepare customized scripts associated with searching indexes and displaying results
G. Add relevant viewers to supported configuration for WWW access in reading rooms
H. Assembly completed

VII. Test and refine

A. Review assembled collection for accuracy and completeness
B. Test links
C. Make any necessary changes
D. Testing completed

VIII. Release Collection

A. Move HTML pages to production area of LCWEB server
B. Provide links to new collection from appropriate points in LCWEB structure
C. Add collection-level MARC record with pointer in 856 field to MUMS
D. Release digital collection to public

IX. Update

Appendix B

Note: The following is from the Library of Congress's website at gopher://marvel.loc.gov/00research/reading.rooms/prints/195_copr.asc

Copyright and Other Restrictions Which Apply to Publication and Other Forms of Distribution of Images: Sources for Information

(P&P Reference Aid #195)

Prints and Photographs Division
Library of Congress
Washington, D.C. 20540-4840

It is not true that all images within the Prints & Photographs Division are in the public domain. Patrons need to be aware of the several kinds of rights which might apply: copyright, donor restrictions, privacy rights, publicity rights, licensing and trademarks. This reference aid leads patrons to information about these rights.

Division policy

As a publicly supported institution the Library generally does not own rights to material in its collections. Therefore, it does not charge permission fees for use of such material and cannot give or deny permission to publish or otherwise distribute material in its collections. It is the patron's obligation to determine and satisfy copyright or other use restrictions (such as donor restrictions, privacy rights, publicity rights, licensing and trademarks) when publishing or otherwise distributing materials found in the Division's collections. The Division's staff will attempt to inform patrons about such restrictions when information is readily available through the Division's records. However, these records are incomplete. The Division provides such information as a service to aid patrons in determining the appropriate use of an image, but *that determination ultimately rests with the patron.* The following sources of information may be used by patrons to determine if restrictions apply.

Sources of information in the Prints & Photographs Division

The Division maintains a Restrictions Notebook which contains information on what is known about restrictions for specific collections. Entries are arranged by collection name and there is an index. The Notebook is not complete. Catalog records will sometimes contain stock phrases to alert patrons to possible restrictions.

Information available from The U.S. Copyright Office, Washington, D.C. 20559

The U.S. Copyright Office provides various circulars with information about the duration of copyright, conducting copyright searches, and "fair use." The following information will lead patrons to parts of the circulars which may be of particular interest:

1) The "fair use" provision of the copyright law allows the making of copies for "...purposes such as criticism, comment, news reporting, teaching, scholarship, or research." (Circular 21, "Reproduction of Copyrighted Works by Educators and Librarians," page 8)

BEST COPY AVAILABLE

2) Published works copyrighted more than 75 years ago are now in the public domain. (Circular 1, "Copyright Basics," page 6)

3) Unpublished works created before January 1, 1978 may be protected by copyright law until the year 2002. (Circular 1,"Copyright Basics," page 6)

4) "Publication is the distribution of copies...to the public by sale or other transfer of ownership, or by rental, lease, or lending. The offering to distribute copies...to a group of persons for purposes of further distribution...constitutes publication." (Circular 1, "Copyright Basics," page 4)

5) Works created after Jan. 1, 1978 are protected for the author/creator's life plus 50 years. (Circular 1, "Copyright Basics," page 6)

6) Works published before March 1, 1989 without a copyright notice risk loss of copyright protection. (Circular 1, "CopyrightBasics," page 4)

7) Works made for hire may be protected by copyright by the employer, not the employee (Circular 1, "Copyright Basics," page 2)

8) Works produced by the U.S. government are not eligible for copyright. (Circular 1, "Copyright Basics," page 5)

9) Copyright searches cannot be considered conclusive (Circular 22, "How to Investigate the Copyright Status of a Work,"page 5)

Patrons who want an image checked for copyright are advised to make a search of the records of the U.S. Copyright Office themselves, hire someone qualified to do so for them, or pay the Copyright Office for a search. Further information is available in Circular 22, "How to Investigate the Copyright Status of a Work,"and from the Search Division of the U.S. Copyright Office (telephone 202-707-6850). As noted above, such searches can not be considered conclusive but at least will show a good faith effort.

The responsibility of determining whether to use an image or not rests with the patron.

Published sources of information on other types of rights

Published books are a valuable source of information about other types of rights and restrictions. The following rights are described by quotes from books available in the Prints & Photographs Division. These are an aid to further reading and are not meant to be inclusive.

The right of privacy

"Photographs of private persons, who are not celebrities or public figures, can be published without their consent only in an editorial context. Even editorial use is perilous, however, if any individual who is depicted is held libeled, held up to ridicule, or misrepresented." *Picture Research: A Practical Guide*. By John Schultz and Barbara Schultz (N.Y.: Van Nostrand, 1991), p. 226. [call number: TR147.S38 1991 P&P]

Right of privacy as applied to public figures

"A statesman, actor, musician, or inventor who asks for and desires public recognition in a large sense surrenders his right of privacy to the public....On the other hand, the privilege of using the picture of a famous person as a subject of new or current interest or for informative purposes does not extend to the commercialization of his personality through forms of treatment distinct

from dissemination of news or information." *Photography and the Law*. By George Chernoff and Hershel Sarbin (NY: AMPHOTO, 1971), pp. 36-37. [call number: KF2042.P45C44 1971 P&P]

Right of publicity

"Unlike the right of privacy, the right of publicity centers on the freedom to exploit and to exclude others from exploiting the commercial value inherent in, for instance, one's own name, voice, signature, photograph, image, or likeness. That prerogative has been described as a person's right to profit from [his or her] general notoriety." *Art Law: the Guide for Collectors, Investors, Dealers, and Artists*. By Ralph E. Lerner and Judith Bresler (NY: Practicing Law Institute, 1989), p. 341. [call number: KF390.A7F45 1986 P&P]

"Courts have determined that celebrities who profit from the commercialization of their names or images will have the right to license them." *Picture Research: A Practical Guide*. By John Schultz and Barbara Schultz (N.Y.: Van Nostrand, 1991), p. 225. [call number: TR147.S38 1991 P&P]

These books offer other information and advice, such as:

Description of "Works for Hire"

"Works considered to be works for hire are an important exception to the general rule that a photographer owns the copyright in a picture he or she has taken. If a photograph was taken by an employee on the job, the law considers the picture a work for hire, and the employer may own the copyright." *The Photographer's Business and Legal Handbook*. By Leonard D. DuBoff (NY: Images Press, Inc. 1989), p. 5 [call number: KF2042.P45D8 1989; being assigned to P&P; xeroxed portions in the P&P Reference file under "Copyright"]

Copying an image from a copyrighted book

"When faced with the requirements of a project and a lack of original photographs of some of the subjects that must be illustrated, publishers will often make a photographic copy of a picture already reproduced in a copyrighted publication. The publisher of the book from which the copy has been made must be contacted and permission requested for the intended use. However, that publisher may not own the copyright to the photograph it has reproduced." *Picture Research: A Practical Guide*. By John Schultz and Barbara Schultz (N.Y.: Van Nostrand, 1991), p. 236. [call number: TR147.S38 1991 P&P]

Practical approaches to unsolvable copyright problems

"...Pictures can fall into a murky area where they may or may not be copyrighted. These situations are perilous to the user, and vexing to the picture researcher or permissions researcher who must try to assure the publisher that he owns the legal right to reproduce. When copyright is unknown or ambiguous, publishers have to make calculated risk decisions...." *Picture Research: A Practical Guide*. By John Schultz and Barbara Schultz (N.Y.: Van Nostrand, 1991), p. 216. [call number: TR147.S38 1991 P&P]

Appendix B (continued)

Copyright infringement and remedies

"The copyright owner may request that the court award actual damages or statutory damages....The amount of statutory damages is decided by the court, within specified limits: no less the $500 and no more than $20,000. The maximum possible recovery is increased to $100,000 if the copyright owner proves that the infringer knew that an illegal act was being committed. The minimum possible recovery is reduced to $200 if the infringer proves ignorance of the fact that the work was copyrighted. The court has the option to award the prevailing party its costs and attorney's fees. As previously noted, statutory damages and attorney's fees may not be awarded in cases where the copyright was not registered prior to infringement, provided such infringement occurred more than three months after the copyrighted work was published." *The Photographer's Business and Legal Handbook*. By Leonard D. DuBoff (NY: Images Press, Inc. 1989), p. 14. [call number: KF2042.P45D8 1989; being assigned to P&P; xeroxed portions in the P&P Reference file under "Copyright"] .